AD 697 343

EVALUATION OF R AND D ORGANIZATIONS

E. M. Glass

Office of the Director of Defense Research and Engineering
Washington, D. C.

31 July 1969

MAM 69-3

EVALUATION OF R&D ORGANIZATIONS

by

E. M. Glass
Assistant Director (Laboratory Management)

July 31, 1969

Management Analysis Report 69-3

Office for Laboratory Management
Office of the Director of Defense Research and Engineering
Washington, D.C. 20301

FOREWORD

One of the 42 problems identified by the Director of Defense Research and Engineering, in his high-priority effort to strengthen the Department of Defense in-house laboratories, concerned the evaluation of research and development organizations.[1] This report describes a peer rating experiment which was an attempt to gain greater insight into the relationship of management and organizational performance with organizational characteristics.

[1] *Problems of the In-House Laboratories and Possible Solutions* (Washington, D.C.: Office of the Director of Defense Research and Engineering, Management Analysis Memorandum (MAM) 66-3, 25 October 1966).

INTRODUCTION

R&D managers have been trying for many years to devise better yardsticks to measure the effectiveness and utility of laboratories. Most managers agree that such techniques are needed, but few can agree on how much appraisals can or should be made. Both "hard" and "soft" scientists have made many attempts to "get a handle" on this question, from physicist and Nobel laureate William Shockley, in his paper, "On Statistics of Individual Variations of Productivity in Research Laboratories," to social scientist Professor Donald C. Pelz, in his study, "Motivations and Working Relations of Scientists and Engineers."

During the period 1966-1968, interest in improving the effectiveness of Federal laboratories heightened. It was during that time that the Congress became interested in the matter of Federal laboratory performers. Both the Subcommittee on Research and Technical Programs (Reuss) of the House Committee on Government Operations and the Subcommittee on Science, Research and Development (Daddario) of the House Committee on Science and Astronautics recommended that greater attention be given to this important area. Six days of hearings were held by the Daddario Subcommittee early in 1968 on the utilization of Federal laboratories.

DADDARIO SUBCOMMITTEE HEARINGS

A good deal of attention was paid to the appraisal of Government laboratories during these hearings. Dr. Donald Hornig, then scientific adviser to the President, described the various techniques used within the Government to appraise laboratories. He pointed out that, to a considerable degree, laboratory evaluations are necessarily qualitative rather than quantitative and involve judgments based on such factors as experience and comparison with good practice elsewhere.

Dr. Allan Astin, Director of the National Bureau of Standards, felt that it was essential to have some mechanism for rating laboratories:

> I think one of my responsibilities is to have techniques for rating the different laboratories within the organization, and where the leadership or management is ineffective, doing something about strengthening it as well as encouraging those that are strong.

Dr. William McLean, Technical Director of the Navy Undersea Warfare Center, believes that a nonsubjective measure of laboratory effectiveness is competition between laboratories and that judgment should be made on the record of their accomplishments. This can only be done, however, by an appraiser who has been successful in the field being evaluated.

Dr. Harold Finger, then Associate Administrator, Office of Organization and Management, National Aeronautics and Space Administration (NASA), described the problem of evaluating the capability and performance of research organizations as a difficult one:

> Perhaps it is one which can be answered only subjectively, since it is difficult to quantify or score so complex and sophisticated an organization as a laboratory.

He believes that NASA experience confirms a widely held view among research administrators that the basic determinant of strength or weakness is the fundamental issue of the value and importance of the organization's purposes, the validity of its objectives and its capability to satisfy those purposes and objectives.

> It would seem hard for a research organization to succeed if its basic purposes and motivations are unimportant. On the other end, the research organization which has an important and significant role has a good start in the critical process of building and holding a research competence.

After deliberating over these and many other thoughts on laboratory evaluation, the Daddario Subcommittee concluded:

> The Subcommittee believes that appraisal of Government Laboratories, both for their scientific and their technical performance, is an essential element of Federal Laboratory administration. . . . It is also apparent to the Subcommittee that Federal principles and procedures for laboratory appraisal are not yet commonly agreed upon nor is there complete agreement that appraisal of scientific productivity and performance is feasible. Nonetheless agency heads responsible for major investments in Government Laboratories cannot wait until the ideal or final method is found. They must go ahead with what is now available and improve it through use.

DOD INTEREST IN APPRAISAL

The Department of Defense (DoD) employs about 60 percent of the civil service engineers and 35 percent of the civil service scientists in the Federal Government, most of whom work in laboratories and test and evaluation activities. In FY 1968, RDT&E (research, development, test and evaluation) obligations for these laboratories were $1.8 billion; about half of this was utilized to support in-house work, the remainder being contracted out. These laboratories employ about 72,000 people, of whom about 28,000 are scientists and engineers. Based upon acquisition costs, the investment in physical plant and equipment is about $2.2 billion.

This is a major investment, and we are anxious to see that it is managed judiciously and effectively and that it is used to solve the pressing problems of the three Military Departments and the six Defense Agencies. This requires appraisals and tough-minded decisions concerning laboratories with poor or marginal capabilities and management encouragement for laboratories of high quality.

Within the DoD many different types of appraisals are regularly made—supervisory evaluations, program evaluations, special appraisals, committee visits, and the natural competition of laboratories for important programs. Most of these techniques are subjective in nature and lack a quantitative basis, particularly for comparisons among laboratories with widely differing missions and technical orientation.

To rectify some of the deficiencies in the current appraisal systems, a comprehensive data base has been developed to provide comparative statistical and trend data on the characteristics and performance of laboratories. Yet this in itself was not considered entirely adequate. It was felt that the utility and significance of these data might be improved if we could relate them somehow to the comparative technical competence or quality of laboratories. In seeking a relatively simple method of ranking laboratories by quality, we elected to use an Apstein-modified Pelz technique.[2]

Apstein-Modified Pelz Technique

From 1960 to 1963, one of the top administrators of the Army's Harry Diamond Laboratories, Dr. Maurice Apstein, an engineer by profession, made a study of "The Role of the Military Laboratory in Defense" as his doctoral thesis at the American University. He attempted to assess the relative competence of a selected group of Government laboratories so that he could examine the relationship of quality with other lab characteristics. He selected and modified a procedure used by Pelz in his study of professional relationships within laboratories.[3]

The technique consists of a series of carefully controlled peer ratings in which a rank order is arrived at by means of a sequence of paired comparisons. The peers, or judges, consist of professional technical people with a substantial degree of industrial, university or Federal laboratory experience, mostly in the management of R&D programs and organizations. Emphasis is placed upon the technical rather than the administrative background of the individual judge so that, in his

[2]Maurice Apstein, "Effectiveness of Military Laboratories as a Function of Contract Activity," *IEEE Transactions on Engineering Management*, EM-12, No. 2, June 1965, pp. 44-50.

[3]D. C. Pelz, *Motivations and Working Relations of Scientists and Engineers* (Ann Arbor, Michigan: Institute of Social Research, University of Michigan, Preliminary Report 4, June 1960).

judgment of a given organization, consideration would be given more to technical competence than to administrative efficiency.

Methodology

The name and location of each DoD laboratory were printed on individual cards. Each judge was given a set of cards with the following instructions:

1) Here is a list of the major R and D installations in the Department of Defense. Please separate them into two piles; those you know and those you do not. For purposes of this exercise, to "know" a laboratory is defined as being sufficiently acquainted with its work to have formed an opinion regarding its technical competence to perform its assigned mission. This opinion need not have been obtained first hand; it may have been formed through reading government reports, technical articles in the open literature, and via inputs from other scientific professionals whose judgment you respect. If there is any question in your mind regarding the validity of your information, place the card in the "unknown" pile.

2) Discard the "unknown" pile and separate the known pile into three groups, GOOD, MEDIUM, and POOR.

3) Now take the GOOD pile and lay the cards in front of you so that they are all in view. Place them in ranking order by selecting first the BEST of the group, then the next best, and so on until you have ranked the entire group. Place this pile aside.

4) Repeat with the MEDIUM pile.

5) Repeat with the BELOW AVERAGE and POOR groups.

6) Combine all the piles in ranking order.

It has been claimed that the survey does not really measure technical competence but instead measures technical reputation. Apstein's assumption was that technical reputation in the scientific community is based upon quality of scientific work. The two terms were therefore considered synonymous.

The conclusions of the Apstein study were as follows:

1) A system of peer ratings and paired comparisons appears to permit the ranking of selected R and D laboratories in order of technical competence.

2) There is a discernible relationship between technical competence and the size of the contract effort administered by the laboratory.

3) A contract program greater than 60 percent of the total laboratory budget appears to represent an administrative burden, which impairs the competence of the technical staff.

4) A laboratory which does an insignificant amount of contract supervision should take special care to insure interaction with the rest of the scientific community.

Current Study

The current study by Mr. Evan D. Anderson, Office for Laboratory Management, ODDR&E, expands upon the original Apstein method in several ways. A tenfold increase in judges is being adopted, and participants represent a much broader spectrum of technical people. The judges selected can be categorized in five groups:

(1) DoD laboratory directors
(2) R&D managers and technical specialists within the Federal Government
(3) Industrial technical specialists, consultants and professionals from nonprofit organizations
(4) Scientists and engineers in academic institutions
(5) Technical specialists in DoD program management and system project offices

This enables us to examine the ratings of laboratories from many different perspectives and points of view. Of particular interest are the attitudes of "customers" of the laboratories. These are the technical personnel in the program management and system project offices who utilize the services of many Defense laboratories and are able to judge them on the basis of fairly specific performance factors.

We have obtained ratings from about 250 judges thus far, and hope ultimately to have about 400 people ranking the laboratories. Selected examples of the preliminary data are described in Table 1 and Figure 1.

The rankings have been normalized by arrangement on a percentage basis, using the decile as the class interval. Examples of high-, medium- and low-ranking laboratories are shown. Six laboratories with staffs (N) of at least 75 were selected for comparison. The distribution shows consistent patterns as one proceeds from the highest rated (laboratory rank No. 1) to the lowest in this subset (laboratory rank No. 56). They are similar to the distributions plotted by Apstein, although he utilized a much smaller sample of judges. The statistical significance of the difference in ratings will be determined when the total rating sample has been analyzed. It is not our intent to develop a precise rank ordering, but to use these rankings as a means of exploring the relationships between technical reputation and the many measurable characteristics of laboratories.

PRELIMINARY RANKING DISTRIBUTION OF SELECTED LABORATORIES

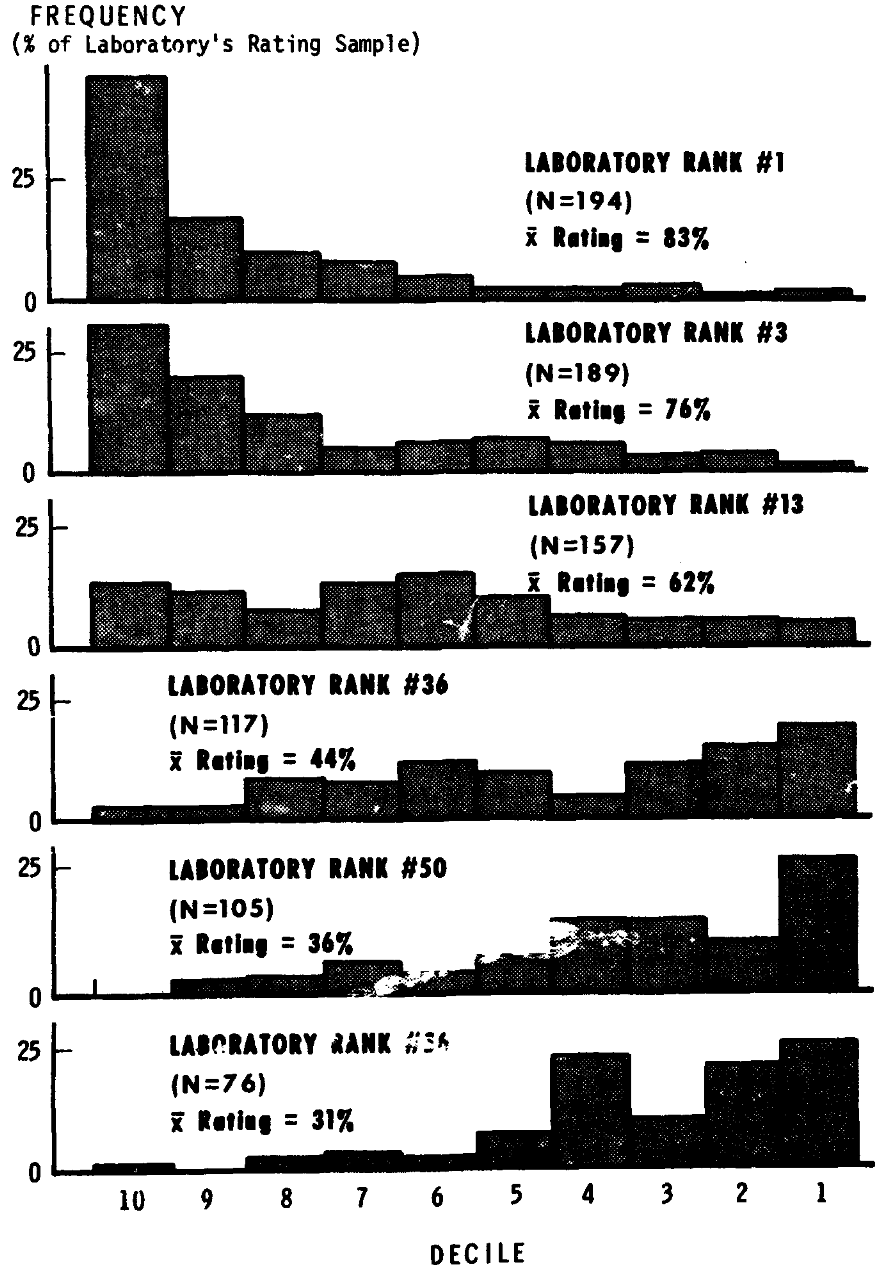

Table 1. RANKING DISTRIBUTION OF SELECTED LABORATORIES

		DECILES (% of Sample)												
		10	9	8	7	6	5	4	3	2	1	Sample N	X Bar Rating %	Median Rating
LABORATORY RANK														
1	%	46.4	17.0	10.3	8.8	5.7	2.6	2.6	3.6	1.0	2.1	194	83	91
Cum.	%	46.4	63.4	73.7	82.5	88.2	90.8	93.4	97.0	98.0	100.2			
3	%	31.2	20.6	12.7	5.8	6.3	7.4	6.3	3.7	4.2	1.6	189	76	84
Cum.	%	31.2	51.8	64.5	70.3	76.6	84.0	90.3	94.0	98.2	99.8			
13	%	14.0	12.1	8.3	14.0	15.3	10.8	7.0	6.4	6.4	5.7	157	62	74
Cum.	%	14.0	26.1	34.4	48.4	63.7	74.5	81.5	87.9	94.3	100.0			
36	%	3.4	3.4	9.4	8.5	12.8	10.3	5.1	12.0	15.4	19.7	117	44	48
Cum.	%	3.4	6.8	16.2	24.7	37.5	47.8	52.9	64.9	80.3	100.0			
50	%	0.0	3.8	4.8	7.6	4.8	8.6	15.2	15.2	11.4	28.6	105	36	30
Cum.	%	0.0	3.8	8.6	16.2	21.0	29.6	44.8	60.0	71.4	100.0			
56	%	1.3	0.0	2.6	3.9	2.6	7.9	23.7	10.5	21.1	26.3	76	31	13
Cum.	%	1.3	1.3	3.9	7.8	10.4	18.3	42.0	52.5	73.6	99.9			
COMPOSITE MILITARY DEPARTMENT RANK														
Service A	%	11.2	11.3	9.8	10.2	10.0	10.2	9.9	9.3	8.6	9.5	2274	57	63
Cum.	%	11.2	22.5	32.3	42.5	52.5	62.7	72.6	81.9	90.5	100.0			
Service B	%	7.8	10.6	11.7	10.1	12.4	10.3	9.5	9.2	8.8	9.5	1363	56	63
Cum.	%	7.8	18.4	30.1	40.1	52.5	62.8	72.3	81.5	90.3	99.8			
Service C	%	7.0	8.2	10.0	9.5	8.9	9.3	10.5	10.9	12.2	13.5	2387	50	53
Cum.	%	7.0	15.2	25.2	34.7	43.6	52.9	63.4	74.3	86.5	100.0			

The Future

Social scientists might argue that this experiment is much too simple a design to yield accurate knowledge and insight into management and organizational performance. Yet we believe it is at present a useful adjunct to other appraisal techniques we currently employ. Hopefully, the kind of organizational research recently described by Likert and Bowers[4] will provide more meaningful answers to the complex relationships among organizational variables.

The first phase of this study will consist of analyses based on extensive data banks already developed on laboratory operations. From that we hope to understand better why certain laboratories are rated high while others are given low ratings. Is it geographical location or location near or distant from a "rich" academic community? Do their positions in the organizational hierarchy matter? Is organizational size or level of funding a factor? Is their orientation toward research or hardware development relevant? What about their relative levels of in-house work versus work contracted out? Are such output indicators as patents, papers, reports, etc., meaningful? Do staff age, advanced degrees, discipline mixes, educational opportunities, salary levels, turnover rates, etc., show important relationships with quality ranking?

We plan to utilize multiple variate analysis to answer these and similar questions to gain deeper insight into the causes of variation between different samples of scientists and engineers. Hopefully, this will provide a better understanding of the similarities and differences of those laboratories rated high in technical reputation and those that receive lower ratings.

The second phase of the study is expected to encompass a much more meaningful attempt to redesign or supplement present criteria measures to increase understanding, reliability and validity. This would extend the range of variables studied to include consideration of reporting and control procedures, management policies, organization structure, professional attitudes, leadership patterns, etc. An analysis of these additional variables would provide more thorough understanding of the management actions and policies that influence laboratory productivity.

The ultimate goal is to give the managers of DoD laboratories a greater insight into research management and organization. It will help them in effectively using data on laboratory properties, performance and their relationships, for purposes of self-evaluation and self-improvement. Finally, we believe it can assist top management in formulating relevant policies and practices that will create the required organizational environment for the DoD laboratories.

[4] R. Likert and D. G. Bowers, "Organizational Theory and Human Resources Accounting," Annual Meeting of the American Psychological Association, San Francisco, California, September 1968.